RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2015.41.108

ScienceAsia 41 (2015): **108–113**

Increasing the discrimination power of a mitochondrial DNA control region by using hypervariable region 2 polymorphisms, as illustrated in Tai populations of Northern Thailand

Chaiwat Kriengchutima[a], Nadchanok Rodrussamee[a], Wibhu Kutanan[b], Jatupol Kampuansai[a,*]

[a] Department of Biology, Faculty of Science, Chiang Mai University 50200 Thailand
[b] Department of Biology, Faculty of Science, Khon Kaen University 40002 Thailand

*Corresponding author, e-mail: Jatupol.k@cmu.ac.th

Received 15 Nov 2014
Accepted 13 Feb 2015

ABSTRACT: Nucleotide sequences of the mitochondrial DNA hypervariable region 2 (HVR2) of five Tai speaking populations, Khuen, Lue, Tai Yai, Yong, and Yuan, living in the upper northern part of Thailand were analysed. Based on a comparison of 268-bp sequences in 124 people, 32 polymorphic sites, and 52 different sequence types were observed. The HVR2 possessed lower polymorphism than HVR1 of the same populations. However, when the two regions were combined, the discrimination power increased and could distinguish most of the Tai populations as well as separate the Tai from the neighbouring Mon-Khmer speaking populations. The Tai showed fewer genetic differences than those of the Mon-Khmer. This might have been a result of the admixture process across the Tai ethnic groups. The increasing of genetic polymorphism and discrimination power of the combined HVR1 and 2 data would enhance the accuracy and efficiency of personal and ethnicity identifications.

KEYWORDS: genetic structure, forensic marker, maternal lineages

INTRODUCTION

Molecular genetics plays an important role in identifying individuals or populations. Traditionally, genetic markers have relied on nuclear DNA, but if the collected samples are in low quantity or degraded, the results may be inadequate. Recently, mitochondrial DNA (mtDNA) variations have been developed to be used as a genetic marker. This has several advantages over using nuclear DNA such as (1) a higher copy number of mtDNA in each cell; and (2) the mtDNA of each person is maternally inherited without chromosomal recombination[1].

The mtDNA sequence generally used as a genetic marker is in the non-coding control region. Since the DNA in this region has no effect on gene expression, changes in the nucleotide sequences are generally observed and thus regarded as highly variable. This mtDNA fragment is known as hypervariable region (HVR), containing three parts: HVR1, HVR2, and HVR3[2]. Although previous research focused mostly on the HVR1, which contains the highest genetic variation, it has been determined that considering HVR1 together with the variation of HVR2 would increase the ability to identify a person. It has also been shown that changes in some HVR2 sequences are specific to some populations[3].

Geographically, the upper northern part of Thailand is made of plains surrounded by multiple mountain ranges with abundant forest. As a result, many ethnic groups migrated to settle in this fertile area. Most of the population living in this region use languages which belongs to the Tai linguistic group[4], including Yuan, Lue, Yong, Tai Yai (Shan), and Khuen. These populations gradually emigrated from Southern China and the borders of Thailand, Laos, and Myanmar 700 years ago[5]. They have diverse cultures, dress, and language dialects. However, genetic studies have shown that these Tai populations were uniquely similar and might have originated from the same ancestors[6,7]. Hence it is necessary to apply highly variable genetic markers to identify each Tai population. This study thus aimed to study the variation of mtDNA HVR2 in the Tai-speaking versus Mon-Khmer-speaking populations, the original ethnic group living in the upper northern part of Thailand since the pre-historical period[8,9].

Table 1 General information of the studied populations.

Code	Ethnicity	Language	Location (village/subdistrict, district, province)	Number of samples
KH	Khuen	Tai	Thung Ruang Thong, Mae Wang, Chiang Mai	24
LU	Lue	Tai	Hea, Pua, Nan	25
YA	Tai Yai	Tai	Mae La Na, Pang Ma Pha, Mae Hong Son	25
YO	Yong	Tai	Ma Kok, Pa Sang, Lamphun	25
YU	Yuan	Tai	Mae Fak Mai, Sansai, Chiang Mai	25
KM	Khamu	Mon-Khmer	Huay Sa Taeng, Thung Chang, Nan	26
LW	Lawa	Mon-Khmer	Dong, Mae La Noi, Mae Hong Son	31
MO	Mon	Mon-Khmer	Nong Du, Pa Sang, Lamphun	14
PA	Paluang	Mon-Khmer	Nor Lae, Fang, Chiang Mai	35
PL	Blang	Mon-Khmer	Pa Yang, Mae Sai, Chiang Rai	16
TN	Htin	Mon-Khmer	Ta Luang, Pua, Nan	29

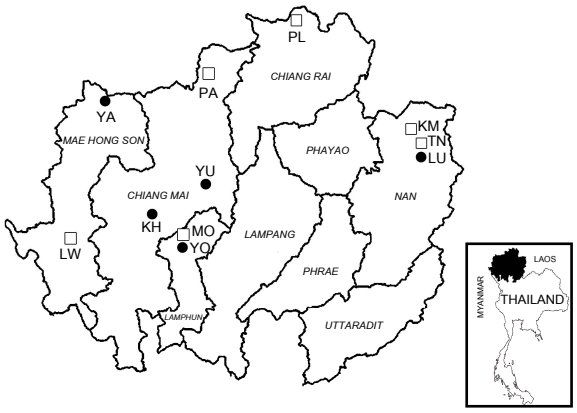

Fig. 1 Geographic distribution of the samples: filled circles, Tai linguistic groups; blank squares, Mon-Khmer linguistic groups.

MATERIALS AND METHODS

The populations used in this study were 124 volunteers from 5 different Tai-speaking populations (Table 1 and Fig. 1). White blood cells of each volunteer were obtained with informed consent from previous mtDNA HVR1 studies[6,10]. All subjects were interviewed regarding their language, personal background, migration, and village history.

Total genomic DNA was extracted from the white blood cell samples according to a standard inorganic salting-out protocol[11]. The quality and quantity of DNA were inspected using a spectrophotometer. The mtDNA control region was amplified using the published primer pairs[2]. The purified PCR products were sequenced for HVR2 at Macrogen Co., Ltd., South Korea.

The mtDNA HVR2 sequence were edited, assembled, and aligned with the revised Cambridge Reference Sequence[12] using SEQSCAPE v2.5 (Applied Biosystems, Foster City, CA). Polymorphic sites and number of haplotype were identified by DNASP v5[13]. Data of mtDNA HVR1 of the same studied Tai individuals[6,10], together with mtDNA HVR1[7] and HVR2 (unpublished data) of the Mon-Khmer speaking populations living in Northern Thailand, were integrated into the analysis. The data were arranged into three sets: (1) HVR1, (2) HVR2, and (3) HVR1 combined with HVR2 (HVR1+2). Haplotype diversity of each data set was calculated by ARLEQUIN 3.5[14]. The discrimination power (D_p) was calculated using the equation $D_p = 1-\sum p^2$ (where p is the frequency of haplotype).

Genetic structure and differences among the populations were analysed by the analysis of molecular variance (AMOVA)[15], using ARLEQUIN 3.5[14]. Variation was divided into three levels: (1) within a population, (2) among populations in the same linguistic group, and (3) between the two linguistic groups (Tai and Mon-Khmer)[4]. Genetic differences between each pair of the populations were examined by pairwise distance using ARLEQUIN 3.5[14].

RESULTS

mtDNA HVR2 sequences, obtained from 124 volunteers of 5 Tai-speaking populations, revealed 52 haplotypes when compared with the Cambridge Reference Sequences positions no. 73–340 (268 bp)[12]. These haplotypes were derived from 32 polymorphic sites, comprising 24 transition, 2 transversion (A95C, A302C), 5 insertion/deletion, and 1 two-variable sites (T146C, T146A). The polymorphic sites most frequently found were A73G (100%), A263G (99.2%), and 318C-insertion (98%). The haplotype frequency found in each population also differed. (Table 2)

The HVR1 data set showed higher haplotype diversity than the HVR2 in every population, except

Table 2 Haplotype of mtDNA HVR2 found in five Tai populations (showing only the variable sites from Cambridge Reference sequences) and number found in each population indicated on the right.

	1111111111222222222222333333 7994455588899000112234446801114 3353601235959047056854793220128O AAAGTCCTAGATTATGAATGAAGAATA––T–A	KH	LU	YA	YO	YU
1	G.....T.....G..C–.C.	1	1			
2	G......C.....G..C–.C.	1			1	
3	G..AC.....G..CC.C.	1				
4	G.....G..C–.–.	1				
5	G.....G.....C.....G..C–.–.	1				
6	G.....G..C–.C.	2	4		4	2
7	G...C.....G..––.C.	2		2		1
8	G.....A.....–G..C–.C.	2			1	
9	G.....G..A....G..CC.C.	2				
10	G.....G.....G..CC.C.	2				
11	G.....G.....C.....G..C–.C.	2				
12	G.....T.....G..CC.C.	2				
13	G....T.....–G..––.C.	2				
14	G....T......C.....G..C–.C.	3				
15	G....T......C.C.....G..C–.C.		1	1		2
16	G.....G..CC.C.		1	1		
17	G......CG.....G..G.CCCCC.		1			
18	G.....A.G..C–.C.		1			
19	G.....G.....G..––.C.		1			
20	G....T......C.CA.....G..––.C.		1			
21	G....TT.....G..CC.C.		1			
22	G.....G.....G..C–.C.		2	1		2
23	G....T......C.C.....G..––.C.		3			
24	G.....–G..––.C.		8	2		6
25	G....T......C.....G..CC.C.			1	1	
26	G....T.....G..C–.C.			1		4
27	G...C.....G..CC.C.			1		
28	G......C.....A.....–G..C–.C.			1		
29	G......C.....G..––.C.			1		
30	G......C.....–G..CC.C.			1		
31	G......C.....–G..C–.C.			1		
32	G.....C.....G..C–.C.			1		
33	G....T.....C.....–G..––.C.			1		
34	G....T....G.C.....G..C–.C.			1		
35	G.....A.....G..C–.C.			2		
36	G.....–G..C–.C.			2		
37	G.....T.....GC.C–.C.			2		
38	GGC.C..C.....G..C–.C.			2		
39	G...CTTC...C.....G..CC.C.				1	
40	G......C.....G.....G..C–.C.				1	
41	G.....AG.....C.....G..––.C.				1	
42	G.....C.....–G..CC.C.				1	
43	G.....G..––.C.				1	
44	G.....A.....G..C–.C.				3	
45	G.....G.....G..––.C.				3	
46	G......C.....–G..––.C.				7	
47	G...A.....C.....G..––.C.					1
48	G...C.....C.....G..––.C.					1
49	G...CT.....C.....G..––.C–					1
50	G.....TC.....G.....G...G..C–.C.					1
51	G....T.....C–.C.					1
52	G..AC.....G..C–.C.					3

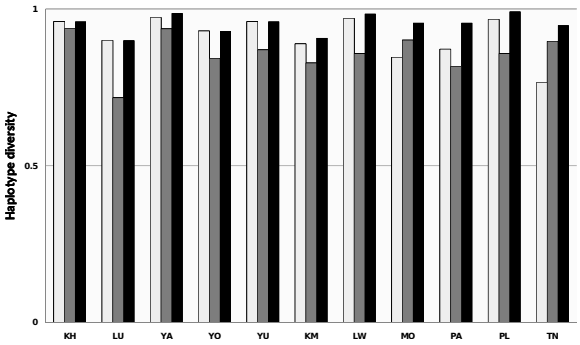

Fig. 2 Haplotype diversity of mitochondrial DNA hypervariable region in each population (white: HVR1, grey: HVR2, black: HVR1+2).

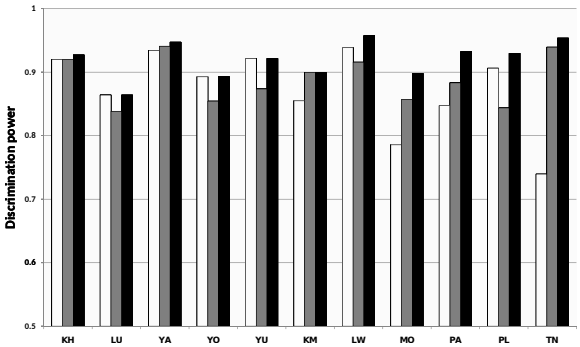

Fig. 3 Discrimination power of mitochondrial DNA hypervariable region in each population (white: HVR1, grey: HVR2, black: HVR1+2).

Mon and Htin. When the two mtDNA regions, HVR1 and HVR2, were combined, the haplotype diversity was higher than 0.9 in every population. The highest HVR1+2 diversity was in the Tai Yai of Pang Ma Pha district, Mae Hong Son province (YA, 0.9941), and the lowest was in the Lue of Pua District, Nan Province (LU, 0.9000) (Fig. 2).

In forensics, genetic markers are usually used to identify individuals through its discrimination power. This study found that the Htin (TN) from Pua District, Nan Province had a low discrimination power compared with other populations, when considered only HVR1. When the HVR2 data was integrated, the discrimination power in the Htin increased significantly (Fig. 3). HVR1+2 dataset showed high power of discrimination in every population. The highest HVR1+2 discrimination power was in the Lawa of Mae La Noi District, Mae Hong Son Province (LW, 0.9573), and the lowest was in the Lue of Pua District, Nan Province (LU, 0.8640) (Fig. 3).

Considering two different linguistic groups, Tai and Mon-Khmer, the HVR1+2 dataset showed higher haplotype diversity and discrimination power in both groups than those from HVR1 or HVR2 datasets alone. Increasing of these two parameters was more distinct in the Mon-Khmer than the Tai group (Table 3). Total HVR1+2 discrimination power of the pooled studied populations was 0.9958; or in other words, the HVR1+2 sequences of two unrelated individuals would be identical by chance at a possibility of 0.0042.

Genetic structure of the populations was con-

Table 3 Haplotype diversity and discrimination power of the Tai and Mon-Khmer linguistic groups.

Linguistic group	Parameter*	Hypervariable region		
		1	2	1+2
Tai	H	0.9870	0.8990	0.9880
	D_p	0.9792	0.9563	0.9813
Mon-Khmer	H	0.9700	0.8930	0.9906
	D_p	0.9862	0.9851	0.9953
Total	H	0.9854	0.9010	0.9940
	D_p	0.9899	0.9841	0.9958

* H = Haplotype diversity; D_p = Discrimination power

Table 4 AMOVA of the Tai and Mon-Khmer linguistic groups based on HVR1+2 variations.

HVR	Lin. gr.†	Within populations		Among popul. within groups		Among groups	
		Var.‡ (%)	Φ_{st}	Var. (%)	Φ_{sc}	Var. (%)	Φ_{ct}
1	T	95.5	0.045*	4.5			
	M	87.9	0.121*	12.1			
	T/M	91.6	0.084*	8.6	0.086*	−0.2	−0.002
2	T	96.7	0.033*	3.3			
	M	92.0	0.080*	8.0			
	T/M	93.7	0.063*	6.1	0.061*	0.2	0.002
1+2	T	95.9	0.042*	4.2			
	M	89.3	0.107*	10.7			
	T/M	92.3	0.077*	7.7	0.077*	−0.1	−0.001

† Linguistic group. T: Tai; M: Mon-Khmer.
‡ Variance.
* statistically significant at $p < 0.05$.

ducted using AMOVA. More than 87% of the genetic variations were found in each population for all three datasets. When comparing the populations who speak the same language, the Mon-Khmer had more genetic differences within their group than the Tai. The genetic difference between the Tai and Mon-Khmer linguistic families was small and no statistical significant differences were observed in any of the datasets (Table 4). Based on the pairwise genetic distance between populations, HVR1+2 dataset could separate 52 pairs (out of 55) of the populations (Table 5), which is an improvement from the 48 pairs if considering HVR1 alone or 36 pairs if considering HVR2 alone (data not shown).

DISCUSSION

In this study, the variations in the mtDNA HVR2 among the ethnic populations of upper northern part of Thailand were investigated. The HVR2 diversities among studied groups were lower than previously reported in populations residing in other regions of Thailand[3]. The most frequently observed variants, A73G and A263G, were not specific to the Tai people of Northern Thailand. They had been found in other Asian populations such as Japanese and Chinese[16–18]. Although haplotype diversity of the HVR2 was lower than HVR1 in all studied populations, the discrimination power increased when both mtDNA regions (HVR1+2) were combined. Increasing of discrimination power is important in ethnicity identification, especially for the closely genetic related populations such as the Tai.

The analysis of mtDNA in HVR1 and 2 compared between the two different linguistic groups—Tai and Mon-Khmer—showed that the discrimination power of the Mon-Khmer (0.9953) was higher than that of the Tai (0.9813), consistent with the AMOVA results that revealed variations within the Mon-Khmer of 11% and the Tai of 4%. These findings indicated that mtDNA marker were more powerful in individual and ethnic identification for the Mon-Khmer than the Tai populations.

The genetic divergence among the Mon-Khmer speaking populations might have been a result from the founder effect[7]. The Mon-Khmer peoples have lived in Northern Thailand since the prehistoric period[8,9], and were subsequently invaded by the Tai from Southern China in the 13th century. After the Tai immigration, the Mon-Khmer people divided into small groups scattered across different mountain ranges. These newly settled Mon-Khmer tended to intermarry within the presently smaller population. As a result, the genetic variance within their population decreased, while the inter-population difference increased. On the other hand, the Tai ethnic group emigrated from the Southern China[5] along the river basins, interacted within and across the ethnic groups in order to create relationships that support their resource management. As a result, the Tai tends to intermarry with other ethnic populations, generating genetic admixture that consequently shaped high genetic variance within their population as well as low inter-population difference. Although the Tai speaking populations were in genetic proximity, the HVR1+2 data could ultimately discriminate most of them. This observation exemplified the potential of using mtDNA variance as a genetic marker in separating the populations living in Northern Thailand.

Table 5 Genetic distance among the Tai and Mon-Khmer populations analysed by mtDNA HVR1+2 (the Tai ethnic groups are in the upper-left square; + significant difference at $p < 0.05$, 10 000 permutations).

	KH	LU	YA	YO	YU	KM	LW	MO	PA	PL	TN
KH		+	−	+	−	+	+	+	+	+	+
LU	0.045		+	+	+	+	+	+	+	+	+
YA	0.024	0.059		+	+	+	+	+	+	+	+
YO	0.047	0.058	0.046		+	−	+	+	+	+	+
YU	0.015	0.037	0.027	0.057		+	+	+	+	+	+
KM	0.077	0.129	0.073	0.037	0.091		+	+	+	+	+
LW	0.051	0.076	0.033	0.077	0.054	0.126		+	+	+	+
MO	0.065	0.104	0.052	0.083	0.068	0.106	0.085		+	+	+
PA	0.046	0.090	0.049	0.115	0.045	0.165	0.034	0.116		+	+
PL	0.065	0.128	0.060	0.109	0.078	0.129	0.061	0.094	0.069		+
TN	0.079	0.113	0.072	0.102	0.052	0.113	0.127	0.114	0.129	0.120	

This study revealed the benefit of using two hypervariable regions (HVR1+2) of mtDNA in personal and ethnic identification, as the results were better than using a single region alone. However, analysing mtDNA HVR1+2 was unable to separate some ethnic populations, e.g., Khuen/Tai Yai, Khuen/Yuan, or Yuan/Khamu. While most of the ethnic populations possessed unique mtDNA haplotypes, some had no particular unique sequence and their haplotypes were shared with other ethnic groups. Thus the genetic structure of the ethnic populations should be investigated prior to using the mtDNA as a genetic marker to identify ethnicity. If cases like the Khuen/Tai Yai, Khuen/Yuan, or Yuan/Khamu occurred, other markers should be used instead of mtDNA.

Acknowledgements: The research team would like to thank all of the volunteers who provided interviews and blood samples for this study. We would also like to thank Professor Bing Su and Xiao-ming Zhang, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, for their technical supports. We would also like to express our appreciation for the advice and assistance from Assoc. Prof. Daoroong Kangwanpong of the Department of Biology, Faculty of Sciences, Chiang Mai University. This study received a grant from the Forensic Science Centre, Chiang Mai University, in 2012.

REFERENCES

1. Cann RL, Stoneking M, Wilson AC (1987) Mitochrondrial DNA and human evolution. *Nature* **325**, 31–6.
2. Hoong LL, Lek KC (2005) Genetic polymorphisms in mitochondrial DNA hypervariable regions I, II and III of the Malaysian population. *Asia Pac J Mol Biol Biotechnol* **13**, 79–85.
3. Mirasena S, Duangchitra S (2012) Mutation rate of mitochondrial DNA at hypervariable region in Phitsanulok province. *Forensic Med J* **4**, 92–105, [in Thai].
4. Lewis MP, Gary FS, Charles DF (2014) *Ethnologue: Languages of the World*, 17th edn, SIL International, Texas.
5. Penth H (2000) *A Brief History of Lanna: Civilizations of North Thailand*, Silkworm Books, Chiang Mai.
6. Kampuansai J, Bertorelle G, Castri L, Nakbunlung S, Seielstad M, Kangwanpong D (2007) Mitochondrial DNA variation of Tai speaking peoples in Northern Thailand. *Sci Asia* **33**, 443–8.
7. Kutanan W, Kampuansai J, Fuselli S, Nakbunlung S, Seielstad M, Bertorelle G, Kangwanpong D (2011) Genetic structure of the Mon-Khmer speaking groups and their affinity to the neighbouring Tai populations in Northern Thailand. *BMC Genet* **12**, 56.
8. Condominas G (1990) *From Lawa to Mon, from Saa' to Thai*, The Australian National Univ Press, Australia.
9. Schliesinger J (2000) *Ethnic Groups of Thailand: Non-Tai-Speaking Peoples*, White Lotus Press, Thailand.
10. Kutanan W, Kampuansai J, Srikummool M, Kangwanpong D (2011) Mitochondrial DNA diversity of the Shan in Mae Hong Son, Thailand. In: *Proceedings of 17th National Genetics Conference*, pp 63–6.
11. Seielstad M, Bekele E, Ibrahim M, Touré A, Traoré M (1999) A view of modern human origins from Y chromosome microsatellite variation. *Genome Res* **9**, 558–67.
12. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147.
13. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–2.

14. Excoffier L, Lischer HEL (2011) ARLEQUIN suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564–7.
15. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–91.
16. Asari M, Umetsu K, Adachi N, Azumi J, Shimizu K, Shiono H (2007) Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. *Leg Med* **9**, 237–40.
17. Irwin JA, Saunier JL, Beh P, Strouss KM, Paintner CD, Parsons TJ (2009) Mitochondrial DNA control region variation in a population sample from Hong Kong, China. *Forensic Sci Int Genet* **3**, e119–25.
18. Liu C, Wang S, Zhao M, Xu Z, Hu Y, Chen F, Zhang R, Gao G, Yu Y, Kong Q (2011) Mitochondrial DNA polymorphisms in Gelao ethnic group residing in Southwest China. *Forensic Sci Int Genet* **5**, e4–10.